Estimateurs

v0 mai 2013 / v0.2 mars 2017 version préliminaire

Ces notes présentent de manière très succincte le cadre, le vocabulaire et quelques exemples de la théorie des estimateurs.

Référence. Le livre de Saporta, le site de la préparation pour l'agrégation de Rennes, ...

1 Cadre

Soit $\mathcal{L}$ un ensemble de mesures de probabilités sur $\mathbb{R}$. Soit $\phi : \mathcal{L} \to \mathbb{R}^d$ une application. Soit $\mu \in \mathcal{L}$. Soit $(X_n)_n$ une suite de v.a.i.i.d. de loi μ.

On dispose d'un ensemble d'observations $x_1, \ldots, x_n$ que l'on suppose convenablement modélisé par $X_1, \ldots, X_n$. L'objectif est d'en inférer des informations sur μ et en particulier (dans le cadre de ces notes) sur $\phi(\mu)$. Plus précisément, il s'agit de proposer une suite d'applications mesurables $T_n : \mathbb{R}^n \to \mathbb{R}^d$ telles que $T_n(X_1, \ldots, X_n)$ (l'estimateur)est proche (en un sens à définir) de $\phi(\mu)$.

Exemples.

— Espérance : $\mathcal{L}$ est l'ensemble des mesures de probabilités telles que $\int |x| d\mu(x)$ est finie et $\phi : \mathcal{L} \to \mathbb{R}$ est définie par

$$\phi(\mu) = \int x d\mu(x).$$

— Variance : $\mathcal{L}$ est l'ensemble des mesures de probabilités telles que $\int x^2 d\mu(x)$ est finie et $\phi : \mathcal{L} \to \mathbb{R}$ est définie par

$$\phi(\mu) = \int x^2 d\mu(x) - \left(\int x d\mu(x) \right)^2.$$

— Cadre Bernoulli : $\mathcal{L} = \{\mathcal{B}(\theta), \theta \in [0,1]\}$ et $\phi : \mathcal{L} \to \mathbb{R}$ est définie par

$$\phi(\mathcal{B}(\theta)) = \theta.$$

— Cadre Gaussien : $\mathcal{L} = \{\mathcal{N}(m, \sigma^2), m \in \mathbb{R} \text{ et } \sigma^2 \geq 0\}$ et $\phi : \mathcal{L} \to \mathbb{R}^2$ est définie par

$$\phi(\mathcal{N}(m, \sigma^2)) = (m, \sigma^2).$$

— Cadre uniforme sur un segment : $\mathcal{L} = \{\mathcal{U}([a,b]), a < b\}$ et $\phi : \mathcal{L} \to \mathbb{R}^2$ est définie par
$$\phi(\mathcal{U}([a,b])) = (a,b).$$

Les cadres Bernoulli, Gaussien et uniforme sont des cadres paramétriques : un élément de $\mathcal{L}$ peut être décrit par la donnée d'un élément de $\mathbb{R}^d$. Par exemple, une loi de Bernoulli $\mathcal{B}(\theta)$ est décrite par le paramètre θ.

Notations. On note souvent et par exemple θ le paramètre que l'on cherche à estimer (l'objet noté $\phi(\mu)$ précédemment) et $\hat{\theta}_n$ voire simplement $\hat{\theta}$ l'estimateur étudié (l'objet noté $T_n(X_1, \dots, X_n)$ précédemment.

Vocabulaire. Considérons un estimateur $\hat{\theta}_n$ d'un paramètre θ.

— Convergent. On dit que l'estimateur est convergent si, pour tout $\mu \in \mathcal{L}$, $\hat{\theta}_n$ converge en probabilité vers θ.

— Fortement convergent. On dit que l'estimateur est fortement convergent si, pour tout $\mu \in \mathcal{L}$, $\hat{\theta}_n$ converge p.s. vers θ.

— Biais. Le biais de l'estimateur est $E(\hat{\theta}_n) - \theta$. C'est une quantité qui dépend a priori de μ et de n.

— Absence de biais. On dit que l'estimateur est sans biais si, pour tout $\mu \in \mathcal{L}$ et tout n, le biais est nul.

— Erreur quadratique moyenne. C'est la quantité
$$E\big((\hat{\theta}_n - \theta)^2\big) = \mathrm{var}(\hat{\theta}_n) + \big(E(\hat{\theta}_n) - \theta\big)^2.$$
Elle dépend a priori de μ et de n. Il est bien sûr préférable que cette quantité soit faible. Il y a toute une théorie sur le sujet (borne inférieure, optimalité, ...).

Intervalles de confiance. Idéalement, on souhaite définir des zones de confiances pour le paramètre que l'on cherche à estimer. Je n'en parle pas dans ces notes. Voir les notes sur les tests.

2 Deux estimateurs incontournables : estimateurs de l'espérance et de la variance

On se place dans le cadre des mesures de probabilité admettant un moment d'ordre un (pour l'espérance) ou deux (pour la variance). On pose
$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$
C'est un estimateur fortement convergent et sans biais de l'espérance. On définit $S_n \geq 0$ par
$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2.$$
C'est un estimateur fortement convergent et sans biais de la variance.

Dans le cas où les X_n sont gaussiennes, la loi de $(\overline{X}_n, S_n)$ est connue (cela peut s'étudier à l'aide du théorème de Cochran) et en particulier

$$\sqrt{n}\,\frac{\overline{X}_n - E(X)}{S_n}$$

suit une loi de Student à $n-1$ degrés de liberté. Dans le cas général,

$$\sqrt{n}\,\frac{\overline{X}_n - E(X)}{S_n}$$

converge en loi vers une loi normale centrée réduite. C'est une conséquence de la LGN, du TCL et du lemme de Slutsky : exercice.

Ces résultats donnent une idée de la vitesse de convergence de l'estimateur. Ils permettent également de construire (avec ou sans cuisine) des intervalles de confiance pour l'espérance : exercice important !

Il y a d'autres estimateurs pour l'espérance et la variance.

— Pour une distribution symétrique et sous des conditions raisonnables (exercice !), l'espérance est égale à la médiane et on peut estimer la médiane à partir de la fonction de répartition empirique (c'est une conséquence du théorème de Glivencko-Cantelli).

— Si on connaît l'espérance m, on peut estimer la variance par

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2.$$

Ce nouvel estimateur est fortement convergent, sans biais et de variance inférieure à celle de S_n^2 (considérer la variance de l'estimateur nécessite un moment d'ordre quatre).

— ...

3 Estimateur du maximum de vraisemblance

Il y a toute une théorie sur le sujet. Dans ces notes on se contente d'introduire les idées et de proposer quelques exemples. On se place dans un cadre paramétrique : $\mathcal{L} = \{\mu_\theta, \theta \in D\}$ où D est un sous-ensemble de $\mathbb{R}^d$.

Si les X_i sont à valeurs dans $\mathbb{Z}$ on s'intéresse à la quantité

$$L_\theta(x_1, \ldots, x_n) = \prod_{i=1}^{n} P_\theta(X = x_i),$$

où les x_i prennent des valeurs entières entières et où $\theta \in D$. On peut remplacer $\mathbb{Z}$ par tout ensemble dénombrable. Si les X_i admettent une densité f_θ par rapport à la mesure de Lebesgue, on s'intéresse à la quantité

$$L_\theta(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

où les x_i sont réels. Autrement dit, on s'intéresse dans tous les cas à la densité de la loi de $(X_1, \dots, X_n)$ par rapport à une certaine mesure de référence.

L'estimateur du maximum de vraisemblance est (quand il existe et est unique) le paramètre $\widehat{\theta}$ (dépendant de $(X_1, \dots, X_n)$) qui maximise $\theta' \mapsto L_{\theta'}(X_1, \dots, X_n)$. Autrement dit et très grossièrement, c'est « le paramètre pour lequel l'observation est la plus probable ».

Regarder ce qu'il se passe dans les cas classiques (Bernoulli, Poisson, Gaussienne, loi uniforme sur un segment, Cauchy...). Il est en particulier intéressant de considérer les deux cas suivants.

— Cauchy. La famille
$$\mu_\theta(dx) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2} dx, \ \theta \in \mathbb{R}.$$
La moyenne empirique est-elle un estimateur pertinent ici ? L'estimateur du maximum de vraisemblance est-il défini ?

— Uniforme. La famille
$$\mu_\theta(dx) = \frac{1}{\theta} 1_{[0,\theta]}(x) dx, \ \theta > 0.$$
Comment se compare ici l'estimateur du maximum de vraisemblance par rapport au double de la moyenne empirique ?